



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada



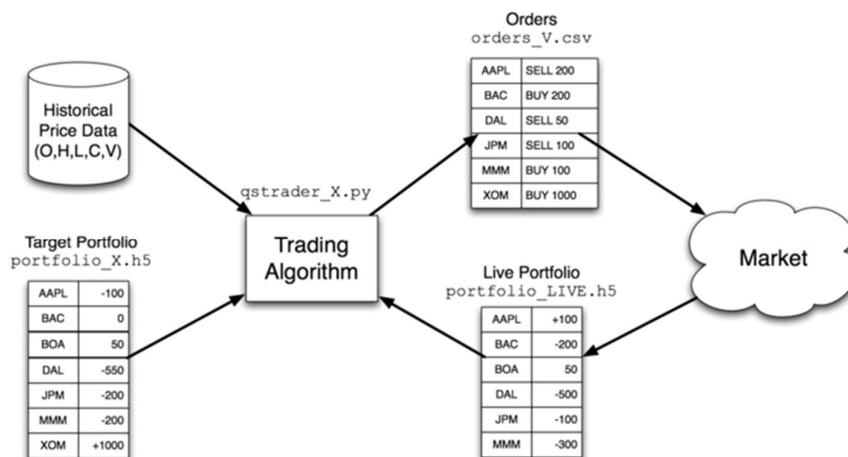
Estudios de eventos

Cómo invertir utilizando algoritmos genéticos

Computational Investing



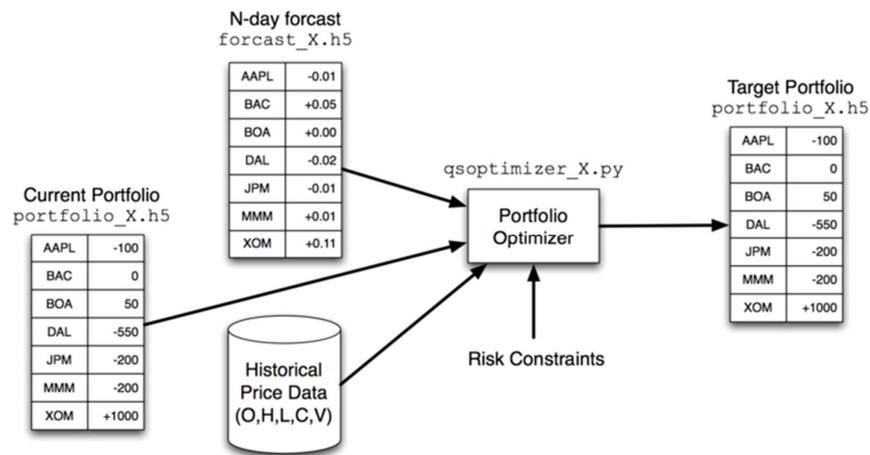
Inside a "quant" shop



Computational Investing



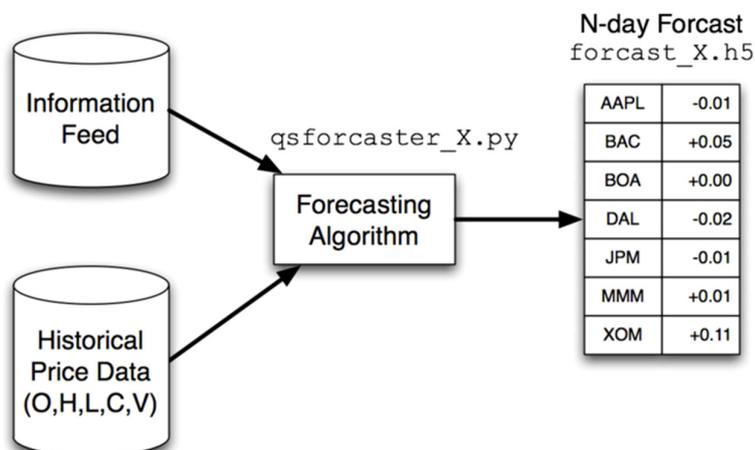
Inside a "quant" shop



Computational Investing



Inside a "quant" shop

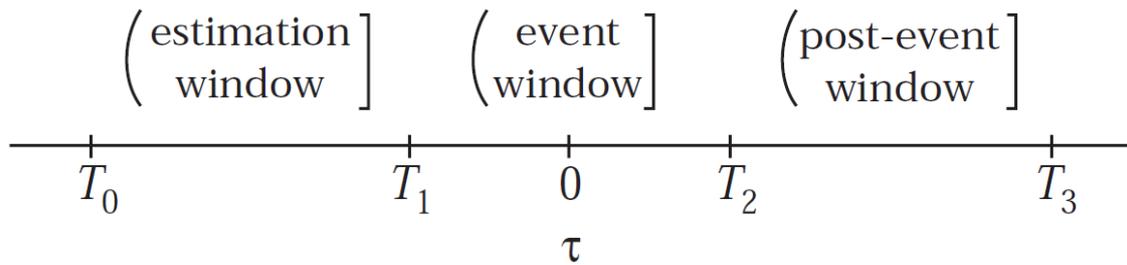


Idea: Event Studies



Time line of an event study

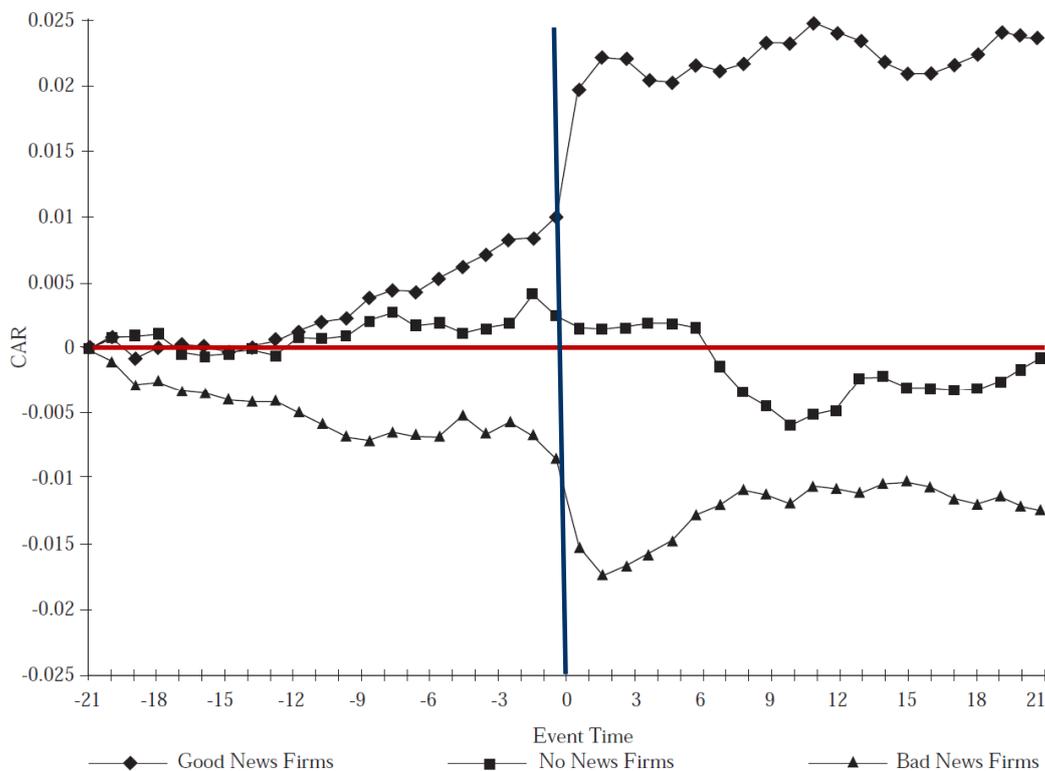
[MacKinlay 1997]



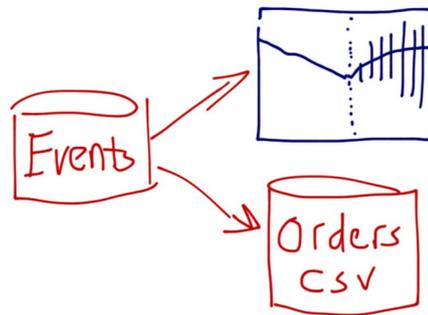
Idea: Event Studies



Cumulative abnormal returns [MacKinlay 1997]



Idea: Event Studies



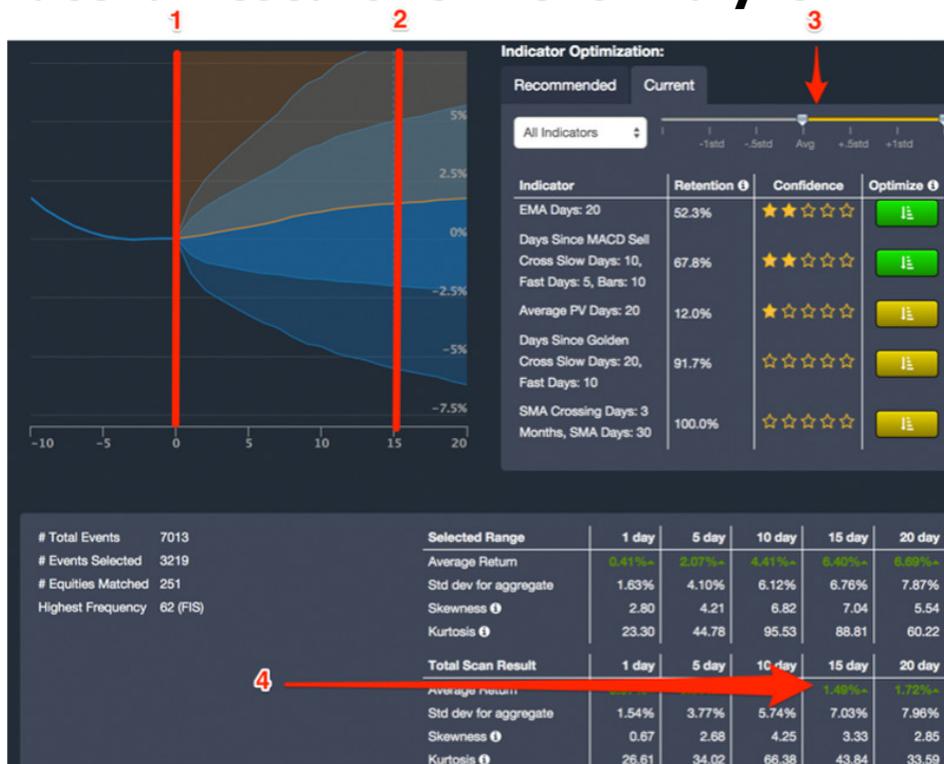
- An event study allows you to test hypotheses.
- A simulator lets you operationalize the transition from hypothesis to strategy...



Idea: Event Studies



Lucena Research's Event Analyzer

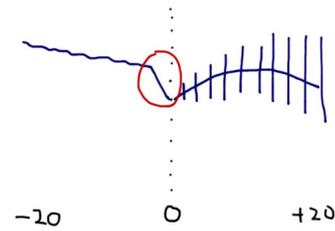


Idea: Event Studies

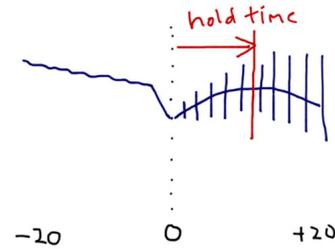


How to Read an Event Study

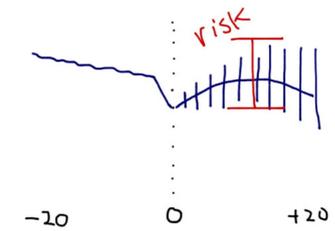
1. Sanity Check



2. Assess: Buy or Sell? Hold Time?



3. Risk? How many opportunities?



Idea: Event Studies



GoToWebinar Viewer

Zoom: 100%

Chrome File Edit View History Bookmarks Window Help

https://quantdesk.lucenaresearch.com/#eventAnalyzer

Back Test GoldDigger, 10 Feb 14 completed

System Messages

- 10/22/14 3:09 PM Intraday Prices Updated
- 9/30/13 12:11 PM Broadcast Messages
- Event Analyzer is live! QuantDesk now supports 5 modules: Forecaster, Optimizer, Hedger, Event Analyzer and Back Tester.

Indicator Optimization:

Indicator	Retention	Confidence	Optimize
Slow Stochastic Days: 20	52.8%	★★★★☆	Optimize
RSI Days: 10	39.6%	★★★★☆	Optimize
Stochastic Days: 5	41.5%	★★★★☆	Optimize
SMA Days: 5	34.9%	★★★★☆	Optimize
Stochastic Days: 20	23.6%	★★★★☆	Optimize
Slow Stochastic Days: 5	40.6%	★★★★☆	Optimize
Slow Stochastic Days: 60	27.4%	★★★★☆	Optimize
SMA Days: 10	50.9%	★★★★☆	Optimize
Alpha: 3 Months	45.3%	★★★★☆	Optimize
EMA Days: 10	32.1%	★★★★☆	Optimize
Stochastic Days: 20	23.6%	★★★★☆	Optimize

1/2 Std Dev 0.00%
First Std Dev 0.00%
Mean 0.00%
Selected Range: 0.00% - Max

	1 day	5 day	10 day	15 day	20 day
Average Return	0.34%	0.31%	0.30%	1.30%	1.34%
Std dev for aggregate	2.08%	3.51%	3.77%	4.45%	4.89%
Skewness	1.72	0.67	0.43	1.13	0.21
Kurtosis	5.63	-0.15	-0.16	0.27	-0.62

Total Scan Result	1 day	5 day	10 day	15 day	20 day
Average Return	0.16%	0.16%	1.84%	2.49%	2.30%
Std dev for aggregate	1.63%	3.51%	4.41%	5.22%	5.86%

Total Events 106
Events Selected 42
Equities Matched 34
Highest Frequency 6 (TWX)

CITRIX Talking: Lucena Research

Idea: Event Studies



GoToWebinar Viewer

Zoom: 100%

Chrome File Edit View History Bookmarks Window Help

https://quantdesk.lucenaresearch.com/#eventAnalyzer

Back Test GoldDigger, 10 Feb 14 completed

System Messages

10/22/14 3:09 PM
Intraday Prices Updated

Broadcast Messages

9/30/13 12:11 PM
Event Analyzer is live! QuantDesk now supports 5 modules: Forecaster, Optimizer, Hedger, Event Analyzer and Back Tester.

Indicator Optimization:

Recommended Current

All Indicators

Indicator	Retention	Confidence	Optimize
Slow Stochastic Days: 20	52.8%	★★★★☆	Optimize
RSI Days: 10	39.6%	★★★★☆	Optimize
Stochastic Days: 5	41.5%	★★★★☆	Optimize
SMA Days: 5	34.9%	★★★★☆	Optimize
Stochastic Days: 20	23.6%	★★★★☆	Optimize
Slow Stochastic Days: 5	40.6%	★★★★☆	Optimize
Slow Stochastic Days: 60	27.4%	★★★★☆	Optimize
SMA Days: 10	50.9%	★★★★☆	Optimize
Alpha: 3 Months	45.3%	★★★★☆	Optimize
EMA Days: 10	32.1%	★★★★☆	Optimize
Stochastic Power: 252		★★★★☆	Optimize

	1 day	5 day	10 day	15 day	20 day
Average Return	0.94%	3.31%	3.98%	7.96%	7.24%
Std dev for aggregate	2.08%	3.51%	3.77%	4.45%	4.89%
Skewness	1.72	0.67	0.43	1.13	0.21
Kurtosis	5.63	-0.15	-0.16	0.27	-0.62
Total Scan Result	1 day	5 day	10 day	15 day	20 day
Average Return	0.98%	0.76%	1.64%	2.49%	2.36%
Std dev for aggregate	1.63%	3.51%	4.41%	5.22%	5.86%

Total Events 106
Events Selected 42
Equities Matched 34
Highest Frequency 6 (TWX)

Selected Range

Average Return

Std dev for aggregate

Skewness

Kurtosis

Total Scan Result

Average Return

Std dev for aggregate

1 day 5 day 10 day 15 day 20 day

0.94% 3.31% 3.98% 7.96% 7.24%

2.08% 3.51% 3.77% 4.45% 4.89%

1.72 0.67 0.43 1.13 0.21

5.63 -0.15 -0.16 0.27 -0.62

0.98% 0.76% 1.64% 2.49% 2.36%

1.63% 3.51% 4.41% 5.22% 5.86%

CITRIX Talking: Lucena Research

Idea: Event Studies



GoToWebinar Viewer

Zoom: 100%

Event Study Backtesting in QuantDesk

LUCENA RESEARCH

Backtest Event: Gap Up

From Jan 9, 2011 To Sep 16, 2014

Performance Summary

Strategy	Strategy	SP500
Total Value	\$11,899K	\$10,892K
Change %	118.47%	89.83%
Sharpe	1.86	1.00
Sortino	2.02	1.25
Max Draw Down	-7.89%	-16.64%
Volatility	0.89%	0.98%
Alpha	0.00	0.00

Transaction Costs

	Strategy	SP500
# Trades	1062	
% Successful Trades	67.08%	
Commission	\$93,906.54	
Slippage	\$458,899.54	
Roundtrip	\$552	
Total Cost	\$552,000.99	

Analysis: Transactions

Risk Return

Industry Correlation

Beta

Transaction Cost Settings

Step Up

Starting Cash \$ 10000000

Date Range 01/09/2011 09/16/2014

ESPTX

Calculate Transaction Costs

Slippage per Trade 1.00%

Short Interest 0.1%

Comps. per Share \$ 0.010

Lucena Research LLC | www.lucenaresearch.com | 404-907-1702

CITRIX Talking: Lucena Research

Problema de búsqueda



Los eventos de interés suceden cuando coinciden varios factores simultáneamente...

¿Qué factores usamos para determinar nuestra estrategia?
350 posibles factores (técnicos y fundamentales),
cada uno con sus parámetros
(umbrales que determinan cuándo se dispara el evento).

Problema combinatorio de búsqueda...
... abordable con un **algoritmo genético**.



Problema de búsqueda



¿Por qué no se utiliza una estrategia de búsqueda exhaustiva, tipo backtracking o branch&bound?

Número de expresiones booleanas con n variables: 2^{2^n}

- 2^n entradas en su tabla de verdad.
- Cada una de ellas puede ser verdadera o falsa.

Variables	Funciones booleanas	Variables	Funciones booleanas
1	4	32	10^{10^9}
2	16	64	$10^{10^{18}}$
4	65536	128	$10^{10^{38}}$
8	10^{77}	256	$10^{10^{76}}$
16	$2 \cdot 10^{19728}$	512	$10^{10^{153}}$



Problema de búsqueda



¿Por qué no se utiliza una estrategia de búsqueda exhaustiva, tipo backtracking o branch&bound?

Si construimos una expresión con k términos elegidos de los n posibles factores:

$$\binom{n}{k} 2^{2^k} = \frac{n!}{k!(n-k)!} 2^{2^k}$$

k	n=k	n=256
1	4	$256 \times 4 = 1024$
2	16	$32640 \times 16 \approx 522240$
4	65536	$174792640 \times 65536 \approx 10^{13}$
8	10^{77}	$10^{77} \times 4 \cdot 10^{14} \approx 4 \cdot 10^{91}$
16	$2 \cdot 10^{19728}$	$2 \cdot 10^{19728} \times 10^{125} \approx 2 \cdot 10^{19753}$



Algoritmos genéticos



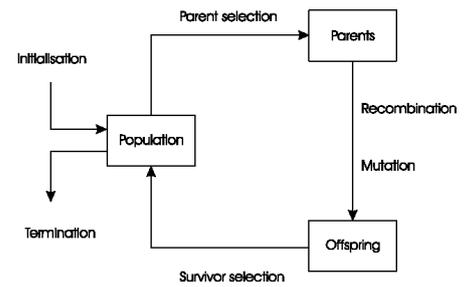
- Se hace evolucionar una población de individuos (cada uno de los cuales representa una posible solución).
- La población se somete a acciones aleatorias semejantes a las de la evolución biológica (mutaciones y recombinaciones genéticas).
- Los individuos se seleccionan de acuerdo con una función de adaptación en función del cual se decide qué individuos sobreviven (los más adaptados) y cuáles son descartados (los menos aptos).





Características esenciales (de cualquier algoritmo evolutivo)

- Algoritmos de tipo “generar y probar” [generate & test].
- Algoritmos estocásticos basados en poblaciones (comportamiento no determinista).
- Operadores de variación (cruce y mutación) crean la diversidad necesaria.
- La selección reduce la diversidad (hacia soluciones de calidad).



Características esenciales

- Inicialización
(generación aleatoria de una población inicial)
- Variación
(operadores de **cruce** y mutación)
- Evaluación
(aptitud [fitness] de cada individuo)
- Selección
(selección **probabilística**)



Algoritmos genéticos

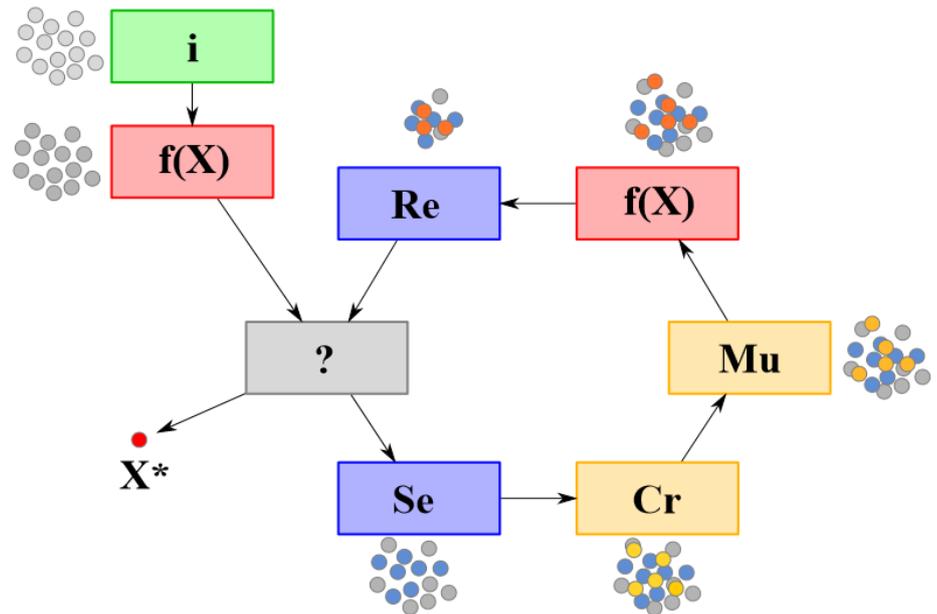


Fases

- Inicialización
- Evaluación

Repetición...

- Selección
- Cruce
- Mutación
- Evaluación
- Reemplazo



Algoritmos genéticos



Algoritmo genético clásico

```
t ← 0
población(t) ← poblaciónInicial
EVALUAR(población(t))
```

```
while not (condición de terminación)
  t ← t + 1
  población(t) ← SELECCIONAR(población(t-1))
  población(t) ← CRUZAR(población(t))
  población(t) ← MUTAR(población(t))
  EVALUAR(población(t))
```

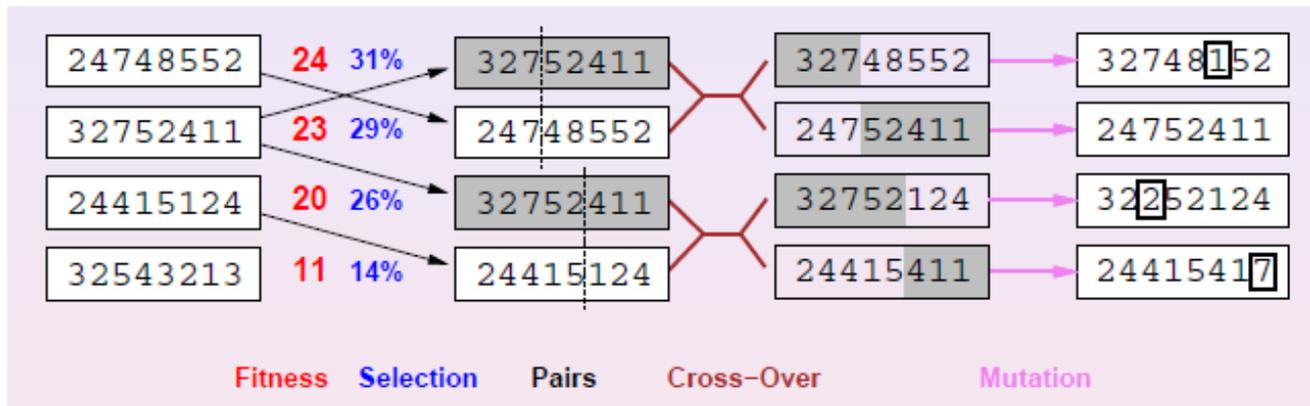
```
return población(t)
```

```
BEGIN
  INITIALISE population with random candidate solutions;
  EVALUATE each candidate;
  REPEAT UNTIL ( TERMINATION CONDITION is satisfied ) DO
    1 SELECT parents;
    2 RECOMBINE pairs of parents;
    3 MUTATE the resulting offspring;
    4 EVALUATE new candidates;
    5 SELECT individuals for the next generation;
  OD
END
```



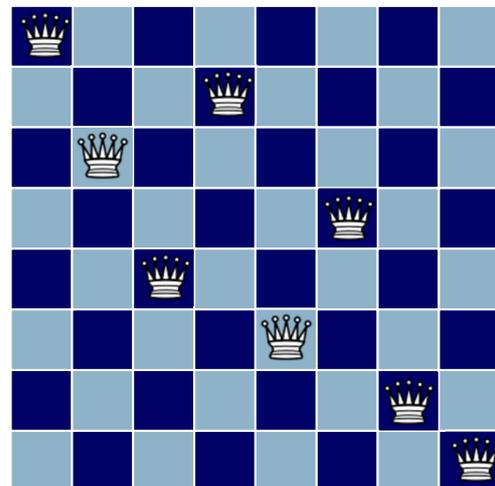


Selección, cruce & mutación



Ejemplo: El problema de las N reinas

- Fenotipo: Configuración del tablero.



- Genotipo: Permutación de enteros.





Ejemplo: El problema de las N reinas

- Función de evaluación:
Número de parejas de reinas que no se atacan.

Definición alternativa como problema de minimización:

- Penalización de una reina =
Número de reinas a las que ataca directamente.
- Penalización del tablero =
Suma de las penalizaciones de todas las reinas.
- Fitness = - penalización del tablero.



Ejemplo: El problema de las N reinas

- Operador de mutación:
Pequeña variación en una permutación.

Ejemplo:

Intercambio de dos posiciones elegidas al azar.

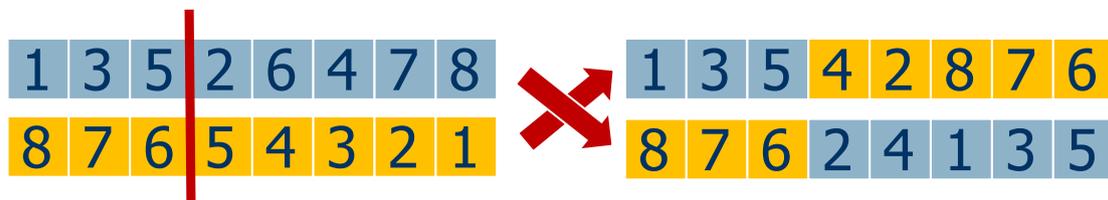
1 3 5 2 6 4 7 8 → 1 3 7 2 6 4 5 8





Ejemplo: El problema de las N reinas

- Operador de cruce:
Combinación de dos permutaciones
 - Elegir un punto de cruce al azar.
 - Copiar la primera parte de ambas.
 - Rellenar la segunda parte usando el otro padre: se añaden valores desde el punto de cruce, en el orden en el que aparecen y saltando los que ya están.



Selección proporcional ("ruleta")

- Idea: Los mejores individuos tienen más posibilidades de reproducirse (de manera proporcional a su fitness).
- Implementación: Ruleta
 - A cada individuo se le asigna una parte proporcional de la ruleta.
 - Se gira la ruleta n veces para seleccionar una población de n individuos.



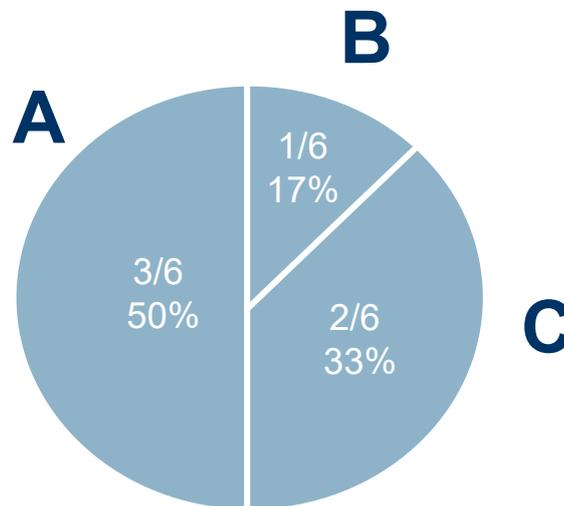


Selección proporcional ("ruleta")

fitness(A)=3

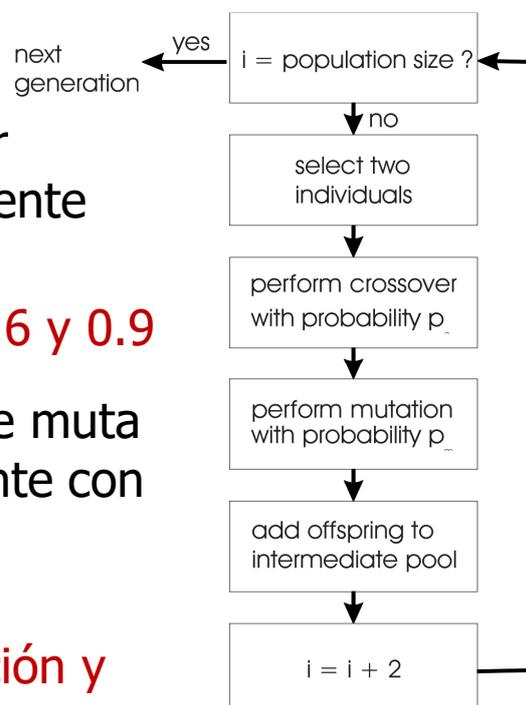
fitness(B)=1

fitness(C)=2



Algoritmo genético estándar

- Seleccionar dos padres.
- Con probabilidad p_c , cruzar los padres (si no, simplemente se copian).
 p_c típicamente entre 0.6 y 0.9
- Para cada descendiente, se muta cada bit independientemente con probabilidad p_m .
 p_m típicamente entre $1/\text{tamaño de la población}$ y $1/\text{longitud del cromosoma}$





Selección de supervivientes (a.k.a. reemplazo)

- La mayoría de los algoritmos evolutivos utilizan una población de tamaño fijo, por lo que hay que definir una forma de crear la siguiente generación a partir del conjunto de individuos disponibles (padres e hijos).
- La selección final suele ser determinística:
 - Basada en el fitness (sólo los mejores sobreviven).
 - Basada en la edad (la generación de descendientes sustituye por completo a sus padres).



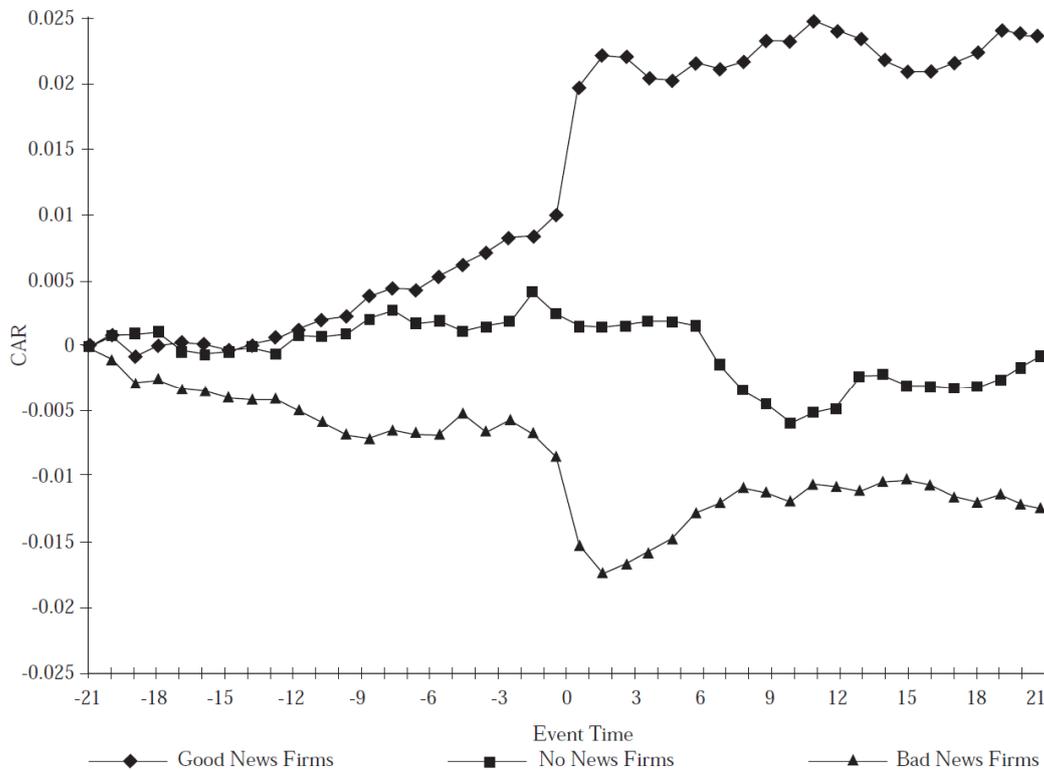
- En ocasiones, se combinan fitness y edad: **elitismo** (el mejor individuo siempre sobrevive, sea padre o hijo).
- Ha sido demostrado que un algoritmo genético requiere elitismo para poder converger al óptimo global.



Günter Rudolph: "Convergence Analysis of Canonical Genetic Algorithms". IEEE Transactions on Neural Networks, 5:96-101, January 1994.



Estudios de eventos



Representación



- Un evento es una lista de factores que hay que considerar.
- Cada factor lleva asociado una pareja de umbrales (mínimo y máximo) que deben sobrepasarse en un día para que se dispare el evento.
- "Nugget" = (factor, min, max)



Implementación



- Selección elitista
(los dos mejores conjuntos de "nuggets" sobreviven).
- Selección proporcional
(padres escogidos en función de su fitness).
- Operador de cruce
(combinación de conjuntos de "nuggets")
- Los umbrales se determinan realizando un proceso de búsqueda particular para cada "nugget".



Fitness



Common performance metrics

- Annual return.
- Risk: Standard deviation of return.
- Risk: Drawdown.
- Reward/Risk: **Sharpe Ratio**.
- Reward/Risk: Sortino Ratio.
- Jensen's Alpha





Sharpe Ratio

$$S = \frac{E[R - R_f]}{\sigma} = \frac{E[R - R_f]}{\sqrt{\text{var}[R - R_f]}}$$

daily_return[i] = (value[i]/value[i-1]) - 1

sharpe = (average(daily_return)/stdev(daily_return)) * sqrt(250)

- **Reward/risk:** Most "important" measure of asset performance.
- How well does the return of an asset compensate the investor for the risk taken? The higher the Sharpe ratio the better.
- When comparing two assets each with the same return, higher Sharpe ratio gives more return for the same risk.

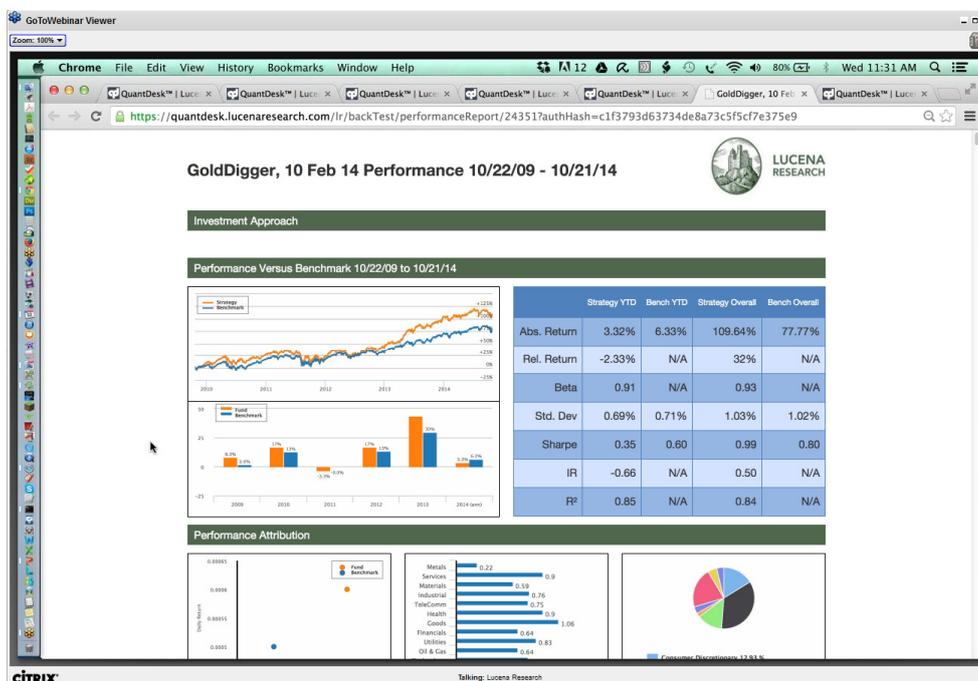


Resultados



GoldDigger

A Genetic Algorithm-based Approach to Market Event Discovery



Lucena Research



**LUCENA
RESEARCH**
Machine Learning Technology
for the Strategic Investor

Lucena brings quantitative analysis and statistical Machine Learning solutions to hedge funds, wealth advisers and advanced individual investors.

Our cloud-based Artificial Intelligence decision support technology enables short term investors and traders to find market opportunities and to reduce risk in their portfolio using technical and fundamental quantitative pattern matching.



Lucena Research CTO



Tucker Balch, Ph.D.
CO-FOUNDER & CTO



<https://www.linkedin.com/in/tuckerbalch/>

Managing Director at J.P. Morgan A.I. Research
Professor at the Georgia Institute of Technology.

Dr. Balch's research centers on Machine Learning. He teaches courses in Artificial Intelligence and Finance. Balch has published over 120 research publications related to Robotics and Machine Learning. His work has been covered by CNN and by New York Times. His graduated students work at Goldman Sachs, Morgan Stanley, Citadel, AQR, and Yahoo! Finance.





Curso online

- Tucker Balch: **Computational Investing**. Coursera.
<https://www.coursera.org/courses?query=investing>

Tutorial

- Michael Kearns:
Algorithmic Trading and Computational Finance,
STOC Tutorial, NYC, May 19th, 2012.
<http://www.cis.upenn.edu/~mkearns/talks/STOCTutorial2012.pdf>

Artículo

- A. Craig MacKinlay: **Event Studies in Economics and Finance**, Journal of Economic Literature, 1997, vol. 35, issue 1, pages 13-39.
<http://www.jstor.org/stable/2729691>

